DOCUMENT RESUME

ED 255 555 TM 850 196

AUTHOR Bolger, Niall
TITLE Gender Difference in Academic Achievement According to Method of Measurement.
PUB DATE Aug 84
NOTE 18p.; Paper presented at the Annual Convention of the American Psychological Association (92nd, Toronto, Ontario, August, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Essay Tests; Foreign Countries; Multiple Choice Tests; National Competency Tests; Secondary Education; *Sex Differences; *Test Bias; *Test Format; Verbal Ability
IDENTIFIERS Ireland

ABSTRACT

This study tests the hypothesis of a gender difference in academic achievement as a function of measurement method. The biasing influence of measurement method on achievement has been recognized. Campbell and Fiske (1959) suggested that a considerable proportion of the variation in test scores may be due to features of the form of test (method) used rather than the individual characteristics (traits) which the test is designed to measure. Using a sample of 15-16-year-old Irish school students, gender differences in achievement were examined for three school subjects (Irish, English, and Mathematics), assessed by two methods (multiple choice test, and written public examination). As has been found in other countries, males performed significantly better than females on multiple choice tests, compared with performance on written examinations. An additional hypothesis that the gender difference would be largest for the languages and smallest for mathematics was not supported. This finding runs contrary to an explanation of this phenomenon in terms of greater verbal skills of females. Alternative explanations are proposed and educational policy implications are discussed. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED255555

Gender difference in academic achievement according to method of measurement

Niall Bolger

Department of Human Development
and Family Studies
Cornell University*
Ithaca, New York 14853
(607) 256-3229

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
N. Bolger
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 850 196

Presented at 92nd Annual Convention of the American Psychological Association at Toronto, Ontario, Canada, August 1984.

*This study was conducted while the author was employed at the Educational Research Centre, St. Patrick's College, Dublin, Ireland. The author is greatly indebted to Dr. Thomas Kellaghan, Director, for encouragement and guidance.

ABSTRACT

This study tests the hypothesis of a gender difference in academic achievement as a function of measurement method. Using a sample of 15-16 year old Irish school students (N=1665), gender differences in achievement were examined for three school subjects (Irish, English, Mathematics), assessed by two methods (multiple-choice test, written public examination). As has been found in other countries, males performed significantly better than females on multiple-choice tests, compared with written examinations. An additional hypothesis that the gender difference would be largest for the languages and smallest for mathematics was not supported. This finding runs contrary to an explanation of this phenomenon in terms of greater verbal skills of females. Alternative explanations are proposed and educational policy implications are discussed.

Gender difference in academic achievement according to method of measurement

The biasing influence of measurement method on research findings has long been recognised, and methodologists have urged that multiple measures of constructs be obtained in order to counteract this potential problem (Campbell and Fiske, 1959; Cook and Campbell, 1979). This issue has acquired relevance in the educational literature due to the finding that males perform relatively better than females on multiple-choice tests compared with written examinations (Dwyer, 1979; Murphy, 1982). Moreover, quasi-experimental effects on public examination scores, have been noted in the United Kingdom, following a change from the use of written questions only to a combination of written and multiple-choice questions. Murphy (1980) provides time-series evidence that following the introduction of a multiple-choice paper into a 1977 public examination in Geography, "the percentage of male candidates obtaining A, B, or C grades became approxiamately 10% higher than the equivalent figure for the female candidates".

We regard these findings as exemplifying the method-trait distinction proposed by Campbell and Fiske (1959). In their now-classic paper, the authors suggest that a considerable proportion of the variation in test scores may be due to features of the form of test (method) used rather than the individual characterictics (traits) which the test is designed to measure. This issue of method variance is central to the present problem, for it appears that substantial gender differences in achievement are attributable to differences in the method by which achievement is measured.

The main explanation offered for these results has been that they reflect a gender difference in verbal ability such that females posess a relative advantage where a written assessment method is used. Evidence for this proposition is provided by Murphy (1982), who found consistent gender differences in achievement according to measurement method (multiple-choice, written test) for a wide range of subjects excepting mathematics. Given the low verbal content required in mathematics examinations, whether written or multiple-choice, this finding supports the verbal hypothesis.

The aims of the present paper are two fold. First, I will test the cross-cultural generalizability of this finding by determining whether a sex difference in achievement, as a function of measurement method, exists in the Irish case. To do so, I will utilise information collected on a cohort of pupils whose academic performance was assessed over a number of years in the 1970s. I will focus on three second-level school subjects (English, Irish, and Mathematics) for which scores on standardised multiple-choice tests and grades on the Intermediate Certificate public examination (taken mid-way through the high school years) are available.

Second, as only one study to date has attempted to explain these findings (e.g., Murphy, 1982), I will test the adequacy of the verbal hypothesis as an explanation of the gender difference (if found) using a more appropriate statistical model (repeated measures ANOVA vs. a sequence of t-tests). Consistent with a concern with the issues identified by Campbell and Fiske (1959), throughout the remainder of the paper I will utilise their terminology by referring to achievement in Irish, English and Mathematics as traits, and by referring to the

types of measurement used, i.e., multiple-choice tests and written public examinations, as methods.

METHOD

Sample

A random sample of 70 high schools, stratified of the basis of gender composition, type, and size, was obtained as part of a large-scale longitudinal study of the effects of standardized testing in the Irish context. The total sample amounted to seventy schools. The subjects in the present study were a cohort of 15-16 year old students who completed multiple-choice tests of Irish, English, and Mathematics in Fall 1975 and sat the state-wide public examinations in June 1976. Complete data were available for 1,565 pupils (773 males and 792 females).

Measures

Multiple choice test scores were obtained using level VI, Form A, of the Drumcomdra series of tests in Irish, English and Mathematics (Educational Research Centre, 1978). These tests, developed shortly before the present data were collected, were designed to assess performance at ages 15-16, when the public examination is usually taken. Public examination results in the three subjects were also obtained.

A difficulty exists in relation to the public examination results in that separate higher level and lower level papers were available in each subject. For purposes of comparison with the multiple-choice test scores, it is necessary to equate higher and lower paper grades. The

method employed in the present study utilises Martin and O'Rourke's transformation (in press). Achievement in each subject was expressed as an eleven-point scale. The scores on the higher and lower papers (in each subject) were mapped onto this scale in the following way: Higher paper, A=11, B=10, C=8, D=7, E=4, F=2, No grade=0; Lower paper, A=9, B=6, C=5, D=3, E=1, F=0, No grade=0.

Design and analysis

For data analytic purposes, a model is required which allows the assessment of gender differences in achievement attributable to both method of assessment and trait assessed. A mixed-model repeated measures analysis of variance (ANOVA) approach, which treats each of the six attainment variables as repeated measures of a single variable, scholastic achievement, is used (Searle, 1971; Winer, 1971). These repeated measures, representing the factors Trait and Method, form the within-subjects design of the ANOVA. The between-subjects design, consists of a single factor, Gender. Note that the fixed components of the mixed-model are Gender, Trait, and Method, whereas the subjects themselves are treated as a random sample.

Confirmation of the hypothesis that males perform better than females on the multiple-choice measures requires that a significant Gender x Method interaction exists. Furthermore, a Gender x Trait interaction is also expected, whereby males, regardless of method of measurement, perform better than females in Mathematics (Maccoby and Jacklin, 1974). Finally, the hypothesis that the method gender-difference is largest for the languages and weakest for Mathematics

presumes a significant three-way interaction.

RESULTS

Raw scores

Mean scores obtained by males and females on the six measures of achievement are presented in Table 1. The group means on multiple-choice Irish are almost identical; a small difference in favour of males exists in the case of multiple-choice English (0.6) units; while a substantial male advantage is evident in multiple-choice Mathematics (9.2 units). The scaled Intermediate Certificate Examination results show a somewhat different picture with the females performing noticably better than males in Irish (0.4 units) and slightly better in English (0.2 units). In Mathematics, once again males show a substantial advantage (0.6 units).

INSERT TABLE 1 ABOUT HERE

Scaled scores

Since the above measures are on different scales, it is not possible from Table 1 to obtain a clear indication of gender differences between sets of measures, e.g., between methods or traits.
To do so, it is necesary to express all measures on a common scale. In this case percentage scores have been used, i.e., in each case 100 is the maximum score attainable. Figure 2 presents these data in the form of the percentage difference between the groups (female score minus male score) on each measure.

INSERT FIGURE 1 ABOUT HERE

Two patterns are discernable in Figure 1. First, there is evidence of gender differences at the level of trait regardless of method of measurement. This is most clear in Mathematics, where males perform substantially better than females. The findings for the languages are less clear, and suggest that method does make a difference. In Irish for instance, the multiple-choice measure indicates no gender difference, whereas the Intermediate Certificate measure indicates that females score 3.5% better than males on average.

The second pattern concerns the consistent way in which the multiple-choice measures, compared to written examinations, favour males, and conversely, the consistent female advantage attributable to Intermediate Certificate measures, relative to multiple-choice. This effect (indicated by arrows) can be seen by comparing the heights of the method columns for each subject. In Irish, the male advantage associated with method of measurement (multiple-choice vs. written examination) is 2.8 percentage points. For English, the male advantage is 2.2 percentage points and for Mathematics, this rises to 3.9 percent.

Multivariate Analysis

Table 2 provides results of the repeated measures ANOVA. The dependent measure was expressed in the form of standard deviate (z) scores. In this way, the main effects of Trait, Method and the Trait x

Method interaction are set to zero, as these are not of interest here.

The hypothesis of a substantial Gender x Method interaction is supported and represents an effect size of approximately 3% of the range of each trait, i.e., on average, females score three percent higher relative to males, when traits are measured by written rather than multiple-choice methods (and vice versa for males). Consistent with the second hypothesis, a large Gender x Trait interaction is also evident. Contrary to expectation, the three-way interaction, whereby the Method by Gender effect varies by Trait, is not significant.

DISCUSSION

This paper tests for the presence of a gender difference in academic achievement according to whether multiple-choice or written examination methods are used. The findings, based on a sample of 15-16 year old Irish students, are that males perform relatively better than females on multiple-choice forms of assessment, compared with written examinations, and vice versa for females. Furthermore, we find that this effect of measurement method is constant across the traits measured (achievement in Irish, English, and Mathematics). Thus, the data indicate the existence of a method-based gender difference, but suggest that this is not attributable to differential verbal skill requirements between the measurement methods used.

Given this lack of support for the verbal hypothesis, several other possibilities warrant consideration, both singly or in combination. First, greater neatness of presentation may explain the superior performance of females in written examinations. Previous research indicates that neatness has a significant effect on achieved scores

(e.g., Briggs, 1980) and is consistent with the finding of a constant gender difference across traits. Second, males may have a greater tendency to guess the answers to multiple-choice questions than females, and thus be more likely, on average, to obtain the correct answer (the subjects in this study were not informed of the advantages of intelligent guessing in multiple-choice tests). These hypotheses have not, to date, been examined.

In the face of these alternative explanations, and given the lack of confirmation of Murphy's (1982) results, we see the next step in this line of research as taking the form of a study which examines the various hypotheses simultaneously. A useful endeavor of this kind would involve the design of multiple-choice and written forms of an examination incorporating identical content, while simultaneously using neatness of scripts (as assessed by raters) and the proportion of non-responses (indexing differences in tendency to guess) as non-experimental independent variables.

A separate issue from that of explaining these findings relates to the educational policy of widespread use of multiple-choice tests in many countries. Assuming the above findings to be correct, then it is likely that the introduction of multiple-choice tests, where this occurs, will tend to improve the performance of males relative to females (as shown in the U.K. by Murphy, 1980). A further implication of these findings is that the introduction of multiple-choice tests will result in changes in the pattern of sex differences such that they will increase in mathematical subjects, i.e., the males advantage in this domain will increase, and decrease in verbal subjects, i.e., the female advantage in this domain will decrease.

In conclusion, the findings reported here are indicative of an effect of method of measurement on the relative performance of males and females at high school level in Ireland. This effect, which confirms the findings of researchers elsewhere, warrants further explanatory research and, in addition, the attention of educational policy makers with regard to the assessment methods used in public examinations.

REFERENCES

Briggs, D. (1980). A study of the influence of handwriting upon grades using examination scripts. Educational Review, 32, 185-193.

Campbell, D. & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cook, T. & Campbell, D. (1979). Quasi-experimentation: Design and analysis issues for field settings. New York: McGraw-Hill.

Dwyer, C. A. (1979). The role of tests and their construction in producing sex-related differences. In M. A. Wittig and A. C. Petersen (Eds.), Sex-related differences in cognitive functioning: Developmental issues. New York: Academic Press.

Educational Research Centre (1978). Drumcondra attainment tests: Level VI, Form A. Dublin: Educational Research Centre, St. Patrick's College.

Maccoby, E. E. & Jacklin, C. N. (1974). The psychology of sex differences. London: Oxford University Press.

Murphy, R. J. (1980). Sex differences in GCE examination entry statistics and success rates. Educational Studies, 6, 169-178.

Murphy, R. J. (1982). Sex differences in objective test performance. British Journal of Educational Psychology, 52, 213-219.

Searle, S. R. (1971). Linear models. New York: Wiley.

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

TABLE 1

Means and standard deviations for multiple-choice and written measures of achievement in Irish, English, and Mathematics: Irish high school students (N=1565).

		MULTIPLE-CHOICE			WRITTEN EXAM		
		Irish	English	Maths	Irish	English	Maths
MALES	Mean	42.0	53.3	61.7	2.9	3.9	3.5
	SD	16.1	15.1	17.4	2.0	1.6	1.9
FEMALES	Mean	42.2	52.7	52.5	3.3	4.1	2.9
	SD	16.6	13.9	16.5	2.0	1.6	1.9

TABLE 2

Repeated measures ANOVA of academic achievement (z-scores) by Trait, Method (within-subjects) and Gender (between-subjects) for a sample of Irish high school students (N=1565).

SOURCE	SS	df	MS	F	p
TOTAL	9382.74	9389			
BETWEEN-SUBJECTS	6591.57	1564			
Gender	21.90	1	21.90	5.21	<.05
Error	6569.67	1563	4.20		
WITHIN-SUBJECTS	2791.17	7825			
Trait	0.00	2	0.00	0.00	N.S.
Gender x Trait	122.59	2	61.29	131.84	<.001
Error	1453.31	3126	0.46		
Method	0.00	1	0.00	0.00	N.S.
Gender x Method	18.20	1	18.20	44.76	<.001
Error	638.30	1563	0.41		
Trait x Method	0.00	2	0.00	0.00	N.S.
Gender x Trait x Method	0.44	2	0.22	1.23	N.S.
Error	558.33	3126	0.18		

FIG. 1 GENDER DIFFERENCES IN ACHIEVEMENT: PERCENTAGE DIFFERENCE BY MEASUREMENT METHOD AND TRAIT